



Sequence analysis of local indicators of spatio-temporal association for evolutionary pattern discovery

Jianing Yu, Hengcai Zhang, Peixiao Wang, Jinzi Wang & Feng Lu

To cite this article: Jianing Yu, Hengcai Zhang, Peixiao Wang, Jinzi Wang & Feng Lu (2025) Sequence analysis of local indicators of spatio-temporal association for evolutionary pattern discovery, *GI Science & Remote Sensing*, 62:1, 2487292, DOI: [10.1080/15481603.2025.2487292](https://doi.org/10.1080/15481603.2025.2487292)

To link to this article: <https://doi.org/10.1080/15481603.2025.2487292>



© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 02 Apr 2025.



Submit your article to this journal 



Article views: 34



View related articles 



View Crossmark data 

Sequence analysis of local indicators of spatio-temporal association for evolutionary pattern discovery

Jianing Yu^{a,b}, Hengcai Zhang^{a,b}, Peixiao Wang^{a,b}, Jinzi Wang^{a,b} and Feng Lu^{a,b}

^aState Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China; ^bCollege of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China

ABSTRACT

The Local Indicators of Spatial Association (LISA) is one of the most widely used methods for identifying local patterns of spatial association in geographical elements. However, the dynamic trends of spatial-temporal (S-T) autocorrelation remain poorly understood, yet capturing these patterns is essential for analyzing the evolution of spatial processes. To fill the gap, we propose a novel S-T LISA methodology to automatically discover co-occurrences LISA subsequences over time by incorporating sequence analysis techniques. First, we extend the classical LISA to a dynamic context, and clarify the definition, properties, and classification of S-T LISA sequences. Second, we adopt an enhanced Hamming distance to quantify the similarity of LISA sequences, followed by hierarchical clustering to group similar LISA sequences. Next, an improved FP-Growth algorithm is applied to identify frequent patterns. Finally, we conduct experiments using grid-scale social media check-in records and city-scale carbon emission data to discover significant evolutionary patterns. The results verified the applicability of the proposed method in both human and physical geography. The proposed approach outperforms traditional S-T cube methods in its ability to automatically capture dynamic, complex, and transient S-T association trends as well as irregular outliers. The integration of sequence analysis with LISA statistics presented in this article provides an effective framework for identifying evolutionary patterns of S-T association.

ARTICLE HISTORY

Received 7 January 2025

Accepted 27 March 2025

KEYWORDS

Spatial-temporal autocorrelation; local indicators of spatial association (LISA); dynamic evolutionary patterns; sequence analysis; hierarchical clustering; frequent pattern mining

1. Introduction

The Local Indicators of Spatial Association (LISA) (Anselin 1995), as one of the most widely used spatial statistical methods, has been extensively applied to identify local patterns of spatial association in various fields, including transportation (Berglund and Karlström 1999), regional economies (Song et al. 2020), disease prevalence (Jesri et al. 2021), and air pollution (Hoffmann et al. 2024). LISA typically considers the localized association as static variables at specific locations. But apparently, the dynamic phenomena are prevalent in geographical analysis. For example, Figure 1(a) shows distribution patterns of microblog check-ins for the adjacent company and shopping center over a week. The number of check-ins not only varies across different locations but also changes dynamically during weekdays and weekends. This highlights the urgent demand for methods which can identify S-T association patterns. As shown in Figure 1(b), the LISA sequences act as genetic encodings of human activity dynamics to quantitatively detect the evolving trends of S-T association, which resemble the role of DNA sequences in geoscience.

The inherent dynamic characteristics of physical world make strong demands for investigating the evolution of spatial structure (Zhang et al. 2024). At present, many studies have tried to incorporate the temporal dimension to the LISA algorithms by extending the spatial weight matrix to S-T weight matrix. For example, S-T distances between observations were calculated to construct S-T weight matrix (Huang, Wu, and Barry 2010; Wu, Li, and Huang 2014). Lee and Li (2017) proposed a method that computed spatial and temporal weight matrices independently and then combined them through multiplication, based on the assumption that space-time effects could be modeled as the product of spatial and temporal effects. Additionally, two S-T proximity structures, the contemporaneous and lagged S-T weight matrices, were proposed and used to extend LISA statistics (Griffith and Paelinck 2018; Wang and Lam 2020). On this basis, Tao, Chen, and Thill (2023) applied these S-T weight matrices to quantify the S-T proximity of flows, proposing S-T flow LISA. The output of all these methods consists of variables calculated through local Moran's I or local Geary's C at

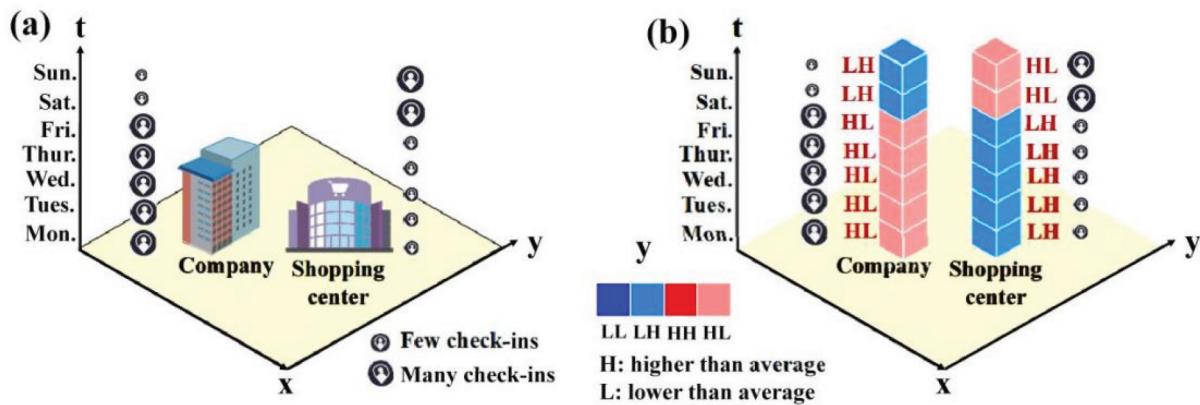


Figure 1. Potential S-T distribution patterns of microblog check-ins for adjacent companies and shopping centers over a week. (a) Schematic diagram of the adjacent companies and shopping center. (b) The LISA sequences proposed in this article.

a specific time and place, generating a series of result maps of the same study area in each timestep. Further comparison of these result maps is necessary to explore the distributional dynamics of spatial processes.

Mapping LISA results at different time periods is one of the most common methods to detect the dynamics of spatial processes. Although this method enables interactive examination of temporal profiles of S-T patterns, visual analysis can be subjective, time-consuming and may overlook certain patterns, especially when dealing with high-dimensional S-T datasets. This limitation puts forward another research direction involving time-series analysis of LISA results to track changes. For instance, the Emerging Hot Spot Analysis and Local Outlier Analysis in ArcGIS detect trends such as new, intensifying, or persistent clusters in a space-time cube (Cheng 2020; Gui et al. 2024; Xu et al. 2022). These methods predefined several patterns according to fixed classification criteria, but they are lack of adaptability and fail to expose undefined hidden patterns. Additionally, the classification of local Moran statistics has been integrated into Markov chain modeling to assess spatial patterns at various temporal intervals (Rey 2010). However, due to the memoryless nature of the Markov chain, this method cannot capture the long-term evolutionary patterns of S-T association. In a related study, a diagram of LISA over time was developed to illustrate the dynamic changes of the spatial value of a particular location in relation to its neighboring regions (Ye and Rey 2013). However, this study only focused on visualization and did not include quantitative calculations of S-T association. Another study measured the spatial autocorrelation

of time-series data based on their similarity (Gao et al. 2019; Guo et al. 2024). However, this method assumes that a location maintains a single S-T autocorrelation pattern over a period, thereby ignoring the dynamic trends in S-T autocorrelation. Research on quantitatively detecting the dynamic evolutionary patterns of LISA at specific local regions is still lacking.

To overcome this problem, we developed a method framework for processing the S-T LISA sequence to explore evolving S-T association patterns. The main contributions of this paper are as follows:

- **Introduction of S-T LISA Sequences:** We propose LISA sequences to model the temporal evolution of local spatial associations, encoding transitions as ordered trajectories. The definition, properties and classification of S-T LISA sequences are clarified.
- **Data-Driven Framework for Evolutionary Pattern Discovery:** We integrate hierarchical clustering and frequent pattern mining to automatically process the S-T LISA sequence, enabling the identification of dominant evolutionary trends and rare anomalies.
- **Validation with Real-World Case Studies:** We demonstrate the framework's robustness through two datasets, carbon emission intensity of 200 cities in China from 2001 to 2020 and over 1.8 million social media check-in records in downtown Beijing in 2017.

The rest of the paper is organized as follows: Section 2 introduces the basic definition and properties of LISA sequences. Section 3 provides a detailed introduction to the architecture of how to identify the dynamic

evolutionary patterns of S-T association. **Section 4** describes two case studies quantifying the temporal shifts of S-T association of carbon emission intensity and human activity patterns respectively. **Section 5** discusses the implications and limitations of this work. **Section 6** concludes the paper and outlines possible future research directions.

2. Preliminary

2.1. S-T LISA

The Local Moran statistic is arguably most widely adopted LISA statistics to identify local spatial clusters (hot spots, cold spots) and outliers. At time t , the local Moran's I of spatial unit i is defined as Equation 1,

$$I_{it} = \frac{v_{it} - \bar{v}}{\sigma^2} \sum_{j=1}^n W_{ij}(v_{jt} - \bar{v}) (i \neq j) \quad (1)$$

where I_{it} represents the local Moran's I statistic of spatial unit i at time t ; σ denotes the standard deviation between observations; v_{it} is the attribute value of region i at time t ; W_{ij} is the S-T weight between unit i and unit j . Considering the influence of temporal nonstationarity mentioned by (Tao and Chen 2023), \bar{v} represents the average of observation value across all time periods, allowing for comparisons of absolute value changes over time. Statistical significance is determined using a conditional permutation method. The specific principles are beyond the scope of this article, and readers may refer to (Anselin 1995) for further details.

The Moran Scatterplot is a graphical expression of LISA, which can intuitively characterize spatial aggregation and dispersion patterns. As shown in formula (2), the x-axis of the Moran scatterplot represents the original variable, while the y-axis represents the spatially lagged variable.

$$x_{it} = v_{it} - \bar{v} \quad (2a)$$

$$y_{it} = \sum_{j=1}^n W_{ij}(v_{jt} - \bar{v}) \quad (2b)$$

All points to the right of the vertical axis have $v_{it} - \bar{v} > 0$ and all points to the left have $v_{it} - \bar{v} < 0$. These values are referred as high and low respectively, in the limited sense of higher or lower than average. Likewise, the values for the spatial lag above and below the horizontal axis can be classified as high

and low. The scatter plot is naturally divided into four quadrants. The upper-right and lower-left quadrants correspond to High-High (HH) and Low-Low (LL) positive spatial autocorrelation respectively. In contrast, the lower-right and upper-left quadrants correspond to High-Low (HL) and Low-High (LH) negative spatial autocorrelation.

2.2. S-T LISA Sequence

2.2.1. Definition of S-T LISA Sequence

For a spatial location s_i , an S-T LISA trajectory $Traj_i$ is obtained by connecting all Moran points corresponding to the region in chronological order. As illustrated in Figure 2(a), the S-T LISA trajectory of s_i is a time-ordered sequence of Moran points associated with this region. The color of the trajectory points in Figure 2(b) represents the spatial quadrants of each trajectory point in $Traj_i$, defining the LISA sequence of location s_i . As previously described, the upper-right and lower-left quadrants correspond to High-High (HH) and Low-Low (LL) cluster types, while the lower-right and upper-left quadrants correspond to High-Low (HL) and Low-High (LH) outlier types. A LISA sequence Seq_i is formally defined as formula(3):

$$Seq_i = \{(q_{i1}, t_1), \dots, (q_{ik}, t_k), \dots, (q_{ip}, t_p)\} \quad (3)$$

where q_{ik} is an elements from the alphabet {NS, HH, LL, LH, HL}, where HH, LL, LH, HL represent statistically significant High-High cluster, Low-Low cluster, Low-High outlier and High-Low outlier, respectively, while NS represents a statistically insignificant case. The parameter P denotes the length of Seq_i .

2.2.2. Characteristics of S-T LISA Sequence

The LISA sequence can be simplified as $\{q_{i1}, \dots, q_{ik}, \dots, q_{ip}\}$. An example LISA sequence, $Seq_i = \langle LL, NS, HL, LH, HH \rangle$, is depicted in Figure 2(b). The LISA sequence exhibits the following key properties:

- Elemental finiteness: the elements of the LISA sequence take values from the set {NS, HH, LL, LH, HL}.
- Time-scale dependence: the length of the LISA sequence is determined by the time scale. Under a uniform time-axis scale, LISA sequences of different locations have the equal length.

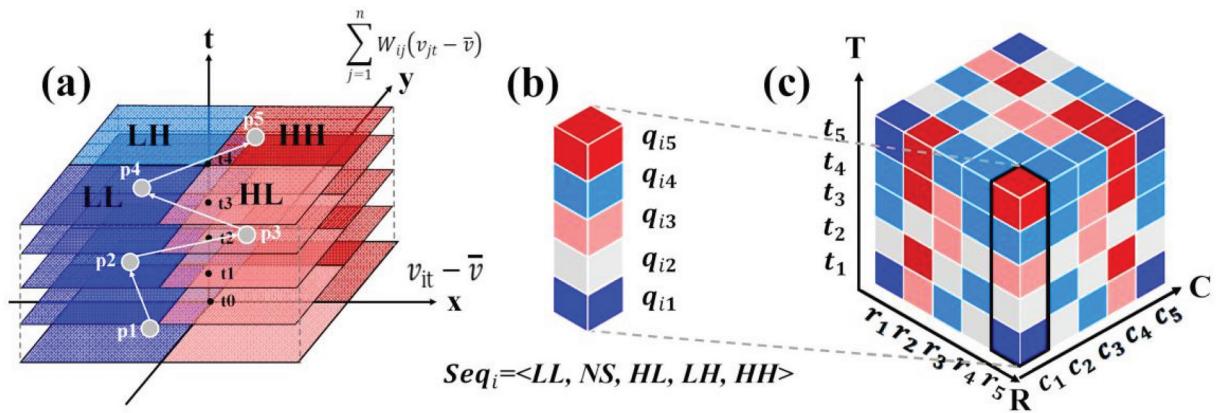


Figure 2. Schematic diagram of the S-T LISA. (a) S-T LISA trajectory. (b) S-T LISA Sequence. (c) S-T LISA Cube ($Q_{5 \times 5 \times 5}$).

- Time-ordered structure: the elements of LISA sequence are arranged in chronological order.

The S-T LISA sequences are categorized as either simple sequences or composite sequences based on the proportion of each mode $Perc_{mode}$. A LISA sequence is considered a simple sequence if it contains only one statistically significant cluster mode, satisfying $Perc_{NS} + Perc_{HH|or|LL|or|LH|or|HL} = 100\%$. Otherwise, it is categorized as a composite sequence with multiple statistically significant cluster modes. Additionally, if $Perc_{HH|or|LL|or|LH|or|HL} > 50\%$, the dominant mode of the composite sequence is determined by the highest percentage.

2.2.3. Collection of S-T LISA sequences

As shown in Figure 2(c), all LISA sequences across the research area collectively construct the S-T LISA cube $Q_{S \times T}$, where S represents space and the T dimension represents time. Each entry q_{st} within this cube corresponds to the association pattern at location s and at time step t. Units associated with the same physical location represent a LISA sequence. Different LISA sequences maintain the same length, dictated by the chosen time scale.

3. Identification of evolutionary patterns of S-T association

This section presents the architecture of the method for identifying evolving S-T association patterns, as illustrated in Figure 3. To begin with, the S-T LISA sequence and S-T LISA cube were

constructed based on the S-T data, which has been introduced in section 2. Then, we improved the hamming distance to measure the similarity of LISA sequences, as detailed in Section 3.1. Based on the generated similarity matrix, the agglomerative hierarchical clustering algorithm is applied to cluster the LISA sequences, as detailed in Section 3.2. Finally, the frequent pattern mining algorithm for interpretation of cluster results is introduced in Section 3.3.

3.1. Similarity measure based on improved Hamming distance

Hamming distance is used to compare the similarity of two sequences of equal length. The elements are compared at corresponding positions in the two sequences. Traditional Hamming distance uses a binary approach: 0 for matches and 1 for mismatches. This simplification treats all mismatches equally, overlooking the semantic difference of mismatches between LISA category. Inspired by the substitution scoring matrix of DNA sequences (Trivedi and Nagarajaram 2020), the modified Hamming distance assigns tiered penalties to LISA category mismatches: 0 for identical elements, 1 for partial mismatches, and 2 for complete mismatches. If one mismatch is statistically insignificant, the average score of 1 is assigned. For any two elements $q, q' \in \{NS, HH, LL, LH, HL\}$, the element-wise distance function $d(q, q')$ is defined as:

$$v_{it} - \bar{v} < 0 \quad (4)$$

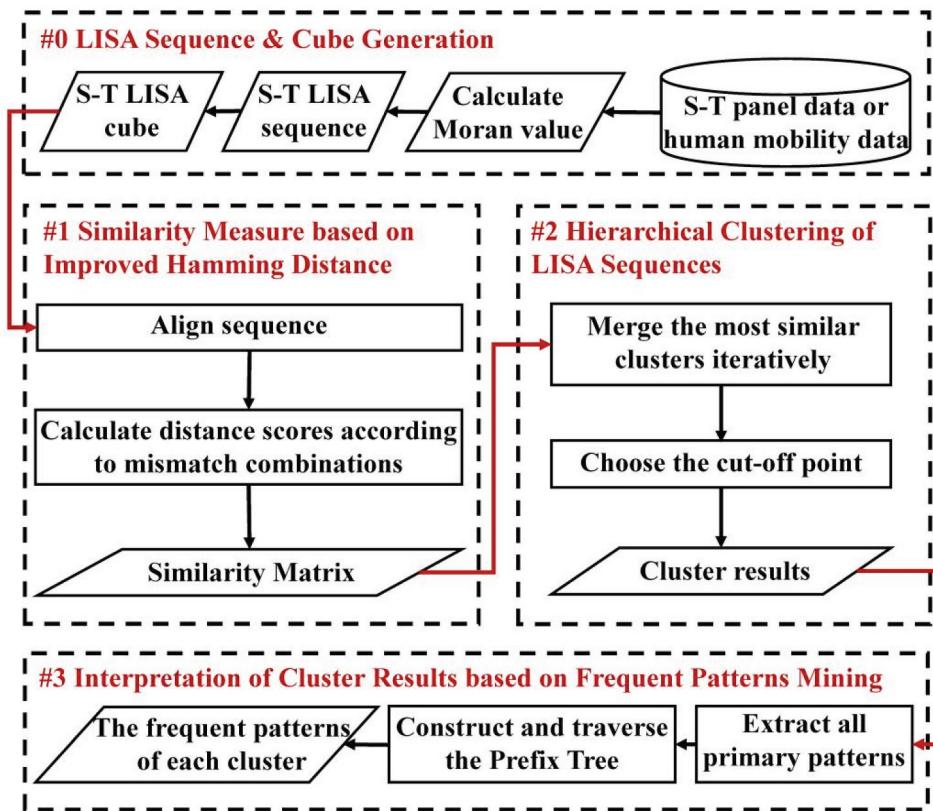


Figure 3. Flowchart of the method proposed in this study.

Then, the similarity between two sequences of equal length is computed by summing the distance scores between elements at each position. To compare the similarity of sequences with varying lengths, the similarity is normalized by dividing it by the sequence length. For two LISA sequences $\text{Seq}_i = (q_{i1}, q_{i2}, \dots, q_{ip})$ and $\text{Seq}_j = (q_{j1}, q_{j2}, \dots, q_{jp})$ of equal length P, the improved Hamming distance is:

$$\text{HamDis}(\text{Seq}_i, \text{Seq}_j) = \frac{1}{P} \sum_{p=1}^P d(q_{ip}, q_{jp}) \quad (5)$$

The comparison between the traditional and modified Hamming distance calculations is depicted in Figure 4. Using the improved Hamming distance, $\{HH, HL, LL, LH, NS\}$ is more similar to $\{HL, HH, NS, LL, LH\}$ than $\{LL, LH, NS, HL, HH\}$. By differentiating mismatch types, our method provides a more precise and nuanced measure of similarity.

3.2. Hierarchical clustering of LISA Sequences

The Agglomerative Hierarchical Clustering algorithm is used to cluster the LISA sequences. Agglomerative

(a)	dis1	dis2	(b)	dis1	dis2
	$\{(1+1+1+1+1)/5 = 1\}$	$\{(1+1+1+1+1)/5 = 1\}$		$\{(1+1+1+1+1)/5 = 1\}$	$\{(2+2+1+2+1)/5 = 1.6\}$

Figure 4. Illustration of the Hamming distance calculation. (a) Traditional Hamming distance. (b) Improved Hamming distance.

Hierarchical Clustering algorithm is an unsupervised machine learning algorithm for data clustering and grouping. The algorithm operates in a bottom-up fashion, commencing with each data point in a cluster and iteratively merging together the most similar clusters until all data points belong to one cluster. Hierarchical clustering naturally supports pairwise distance matrices used in our framework, making it more suitable for our custom metric than alternative methods like k-means, which relies on Euclidean distances, or DBSCAN, which is density-based. Notably, hierarchical clustering does not require a predefined number of clusters, making it ideal for S-T sequences that exhibit complex and evolving behaviors, where the number of meaningful clusters is not known in advance. Additionally, it offers a visual representation of the clustering outcomes through a dendrogram, offering flexibility for a wide range of datasets and application scenarios.

For the LISA sequence set $\{Seq_1, \dots, Seq_i, \dots, Seq_n\}$, the similarity of the LISA sequences is measured by the improved Hamming Distance. The algorithm of the Agglomerative Hierarchical Clustering is as Algorithm 1. Average linkage is selected to calculate the average distance between all pairs of points in two clusters, which helps generate more balanced and representative clusters by minimizing the influence of outliers. The number of clusters is determined based on the dendrogram visualization and the elbow method.

Algorithm 1 Hierarchical Agglomerative Clustering

```

Input: LISA sequence set  $\{Seq_1, \dots, Seq_i, \dots, Seq_n\}$  LS;
Output: A set of LISA sequence clusters C;
1: Create an active set A =  $\emptyset$ ;
2: for all  $i = 1, 2, \dots$ , Iter do
3:   Add  $Seq_i$  as its own cluster:  $A = A \cup \{Seq_i\}$ ;
4: end for
5: Create a cluster set C = A;
6: while  $|A| > 1$  do
7:    $d(C_k, C_l) = \sum_{m=0}^{|C_k|} \sum_{n=0}^{|C_l|} \frac{\text{hamDis}(C_k[m], C_l[n])}{|C_k| \times |C_l|}$ ;
8:    $C_k^*, C_l^* = \text{argmin}(d(C_k, C_l))$ ;
9:   Delete  $C_k^*$ ,  $C_l^*$  from A, add  $\{C_k^*, C_l^*\}$  to A and C;
10: end while
11: return C.

```

3.3. Interpretation of cluster results based on frequent pattern mining

Inspired by FBPM Algorithm (Chen and Liu 2011), which was originated in biological sequences frequent pattern mining, we employed FBPM to discover most commonly LISA subsequences in each cluster,

facilitating the interpretation of pattern characteristics.

The FBPM algorithm is an enhanced version of the FP-Growth algorithm, designed to meet the specific needs of mining sequences. Unlike standard FP-Growth, which processes transactional data consisting of unordered itemsets, the LISA sequence requires a time-ordered representation. To address this, the concept of primary pattern, which is a specific substring within a LISA sequence, is introduced to construct a prefix tree. For the LISA sequence $Seq = \{q_1, \dots, q_i, \dots, q_n\}$, and for any mode $x \in \{NS, HH, LL, LH, HL\}$, if $q_{i_1} = q_{i_2} = \dots = q_{i_m} = x$, $i_1 < i_2 < \dots < i_m$, substring $\{q_{i_k}, q_{i_k+1}, \dots, q_{i_{k+1}-1}\}$ is the k^{th} primary pattern of Seq with respect to x , and denoted as $S_x(k)$. Let's take a LISA sequence of length 10 $\{HH, NS, HL, HH, HH, NS, HL, LH, HH, LL\}$ as an example. Focusing on mode HH, the elements at positions 1, 4, 5, and 9 are HH. The first primary pattern, $S_{HH}(1)$, starts from the first HH and includes all elements up to (but not including) the next HH, forming $S_{HH}(1) = \{HH, NS, HL\}$. By analogy, $S_{HH}(2) = \{HH\}$, $S_{HH}(3) = \{HH, NS, HL, LH\}$,

$$S_{HH}(4) = \{HH, LL\}$$

Then, in order to optimize search efficiency, a candidate primary patterns table is constructed, where modes are arranged in alphabetical order as $\{NS, HH, LL, LH, HL\}$. The primary patterns table of $\{HH, NS, HL, HH, HH, NS, HL, LH, HH, LL\}$ could be easily derived, as shown in Tab.1. The loc column denotes the beginning position of S_x in Seq .

Next step, we build a Prefix rooted tree of primary patterns, according to the candidate primary patterns table. To preserve both sequence order and positional information, the edges of prefix tree represent ordered transitions between LISA states, while the nodes store the location within the sequence. The Prefix Tree is denoted as T, where LISA subsequences are modeled as edges connecting the root to every leaf node, corresponding to a specific primary pattern. The tree is constructed recursively, as outlined in Algorithm 2, starting from the root and expanding layer by layer. Based on the primary patterns in Tab.1, the constructed prefix tree is shown in Figure 5(a). Each node contains a set $\{n_1, n_2, \dots, n_k\}$, representing the starting

Table 2. Computational complexity and parameters of algorithms.

Algorithm	Algorithm Complexity	Parameter	Rationality
Constructing S-T LISA Sequences	$O(P \cdot S \cdot T)$ P: permutations, S: spatial units, T: time steps	Spatial Neighborhood (Contiguity edges corners) Temporal Neighborhood (1 time-step)	To capture both immediate spatial and temporal associations
Hierarchical Clustering	$O(N^3)$ N: the number of LISA sequences	Linkage Criterion (Average linkage) Number of Clusters (6,5)	To reduce outlier influence
FBPM Algorithm	$O(N \times (L + F))$ L: LISA sequence length, F: total lengths of all the frequent patterns	Minimum Support (2) Max Subsequence Length (1 to 3)	Dendrogram Visualization and Elbow Method (Figures 8 and 12) To balance noise reduction and pattern discovery To balance interpretability and computational cost

4. Case study

4.1. Evolving S-T association patterns of human activities

Investigating how the S-T autocorrelation structure of human activities changes over time enables a quantitative comprehension of human mobility (Ma et al. 2020). The availability of vast amounts of geo-tagged check-in data from social media platforms has provided a new perspective for investigating individuals' daily activities (Liu et al. 2022). This case study used microblog check-in records to quantify the dynamic S-T association patterns of human activities.

4.1.1. Study area and datasets

A dataset of 1,851,602 check-in records from Sina microblog within downtown Beijing from 2017/01/01 to 2018/01/06 was used for this investigation. The spatial resolution is set to 500 m and the time interval is set to 1 week. A 500-meter spatial resolution is suitable for capturing neighborhood-level dynamics while maintaining computational feasibility. With 1.8 million check-in records, this resolution ensures a sufficient number of data points per grid cell. Furthermore, considering the typical 100-meter accuracy of social media geotags, aggregating the data into 500-meter grids mitigates positional noise while retaining significant spatial patterns for analysis. The one-week temporal resolution effectively captures weekly human activity trends over one year while minimizing noise from short-term fluctuations. The microblog check-in points were aggregated according to the spatio-temporal units to generate a check-in data cube of $76 \times 77 \times 53$. Figure 6(b) shows the

number of check-in points. To further validate the effectiveness of the proposed method, the relationship between evolutionary patterns of check-ins and urban land use was investigated with the EULUC-China dataset, developed by Tsinghua University. The dataset assigns five primary function tags and twelve subcategories to each parcel. This study focused on the five main functional labels from the dataset, namely residential area, commercial area, industrial area, transportation facility, and public management and service area.

4.1.2. LISA sequence analysis of human activities

Figure 7(a) displays the S-T LISA cube of the study area. The LISA sequences at the marked position in Figure 7(a) were selected for visualization, and the proportion of the five modes in each sequence was counted, as shown in Figure 7(b). Among a total of 4594 LISA sequences within downtown Beijing, there were 1683 simple sequences and 2911 composite sequences. Simple sequences comprised 36.6% of the total sequences, with the LL mode being the dominant type (1392 sequences), followed by the HH mode (253 sequences). Meanwhile, composite sequences constituted 63.4% of the total sequences, with 1025 sequences having dominant types, while the remaining 1886 sequences exhibited various patterns without clear dominant types. These findings indicate that the study area exhibits dynamic and complex S-T autocorrelation patterns, necessitating further exploration and analysis.

The distance between S-T LISA sequences was calculated using the enhanced Hamming distance introduced in Section 3.1 of this paper. A greater Hamming distance indicates lower similarity between

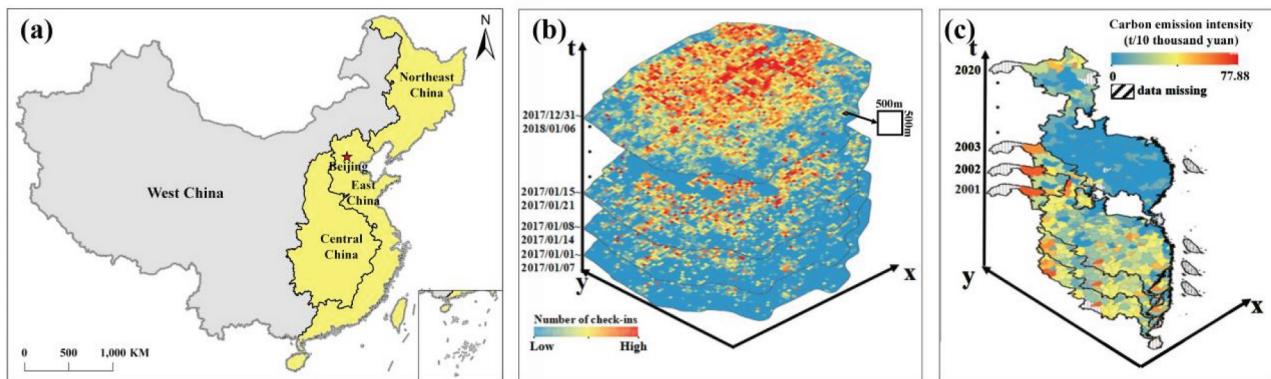


Figure 6. Study area. (a) Map of China highlighted with the study area. (b) Study area of the first case study. (c) Study area of the second case study.

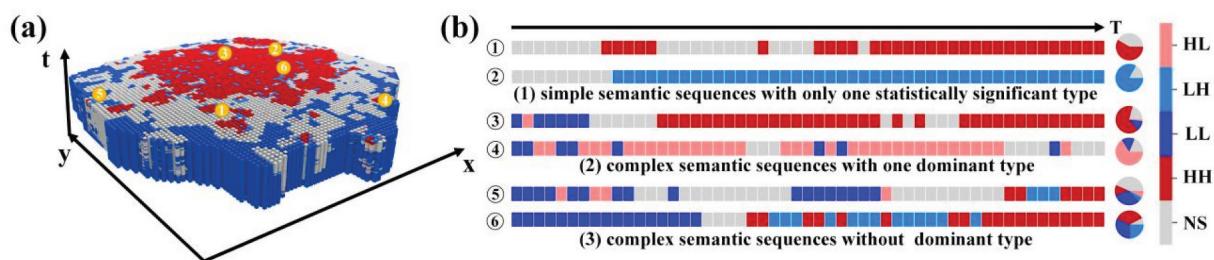


Figure 7. Results of S-T LISA cube and typical LISA sequences of check-in records. (a) S-T LISA cube of the study area. (b) Typical LISA Sequence.

sequences. Figure 8(a) illustrates the histogram of the improved Hamming distance among the sequences, revealing a negatively skewed distribution. This skewness indicates substantial variations in the evolutionary patterns of different LISA sequences. Based on the computed Hamming distance matrix, a hierarchical clustering analysis was conducted. According to the elbow diagram (Figure 8(b)) and dendrogram visualization (Figure 8(c)), the optimal number of clusters was chosen to be 6.

To evaluate the effectiveness of the proposed approach, it is compared with the Local Outlier Analysis (LOA) based on the S-T cube model in ArcGIS Pro. The LOA method is designed to detect local outliers by analyzing the spatial and temporal distribution of data. As shown in Figure 9(a), the LOA method divides patterns into either a single-mode cluster or multiple types. Any patterns exhibiting temporal changes are classified as multiple types, regardless of the nature of the variation. This results in an over-

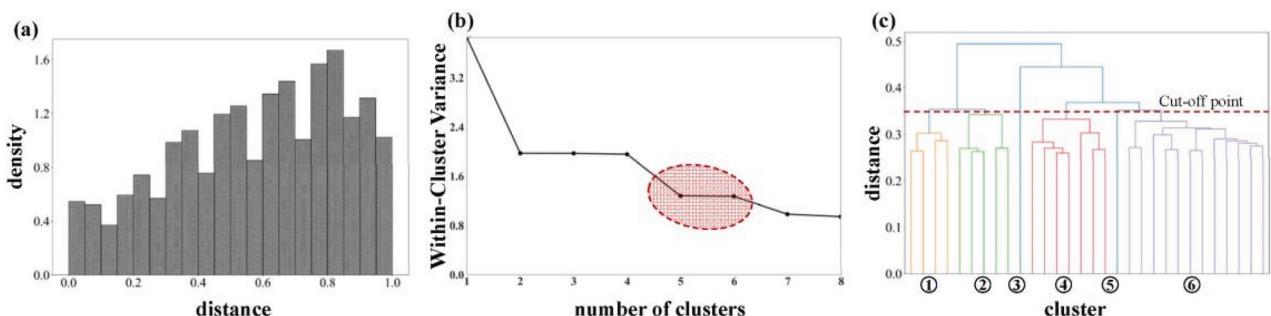


Figure 8. Hierarchical clustering results of check-in records. (a) Histogram of Hamming distances. (b) Elbow plot for determining the optimal number of clusters. (c) Dendrogram of hierarchical clustering with a cut-off point indicated.

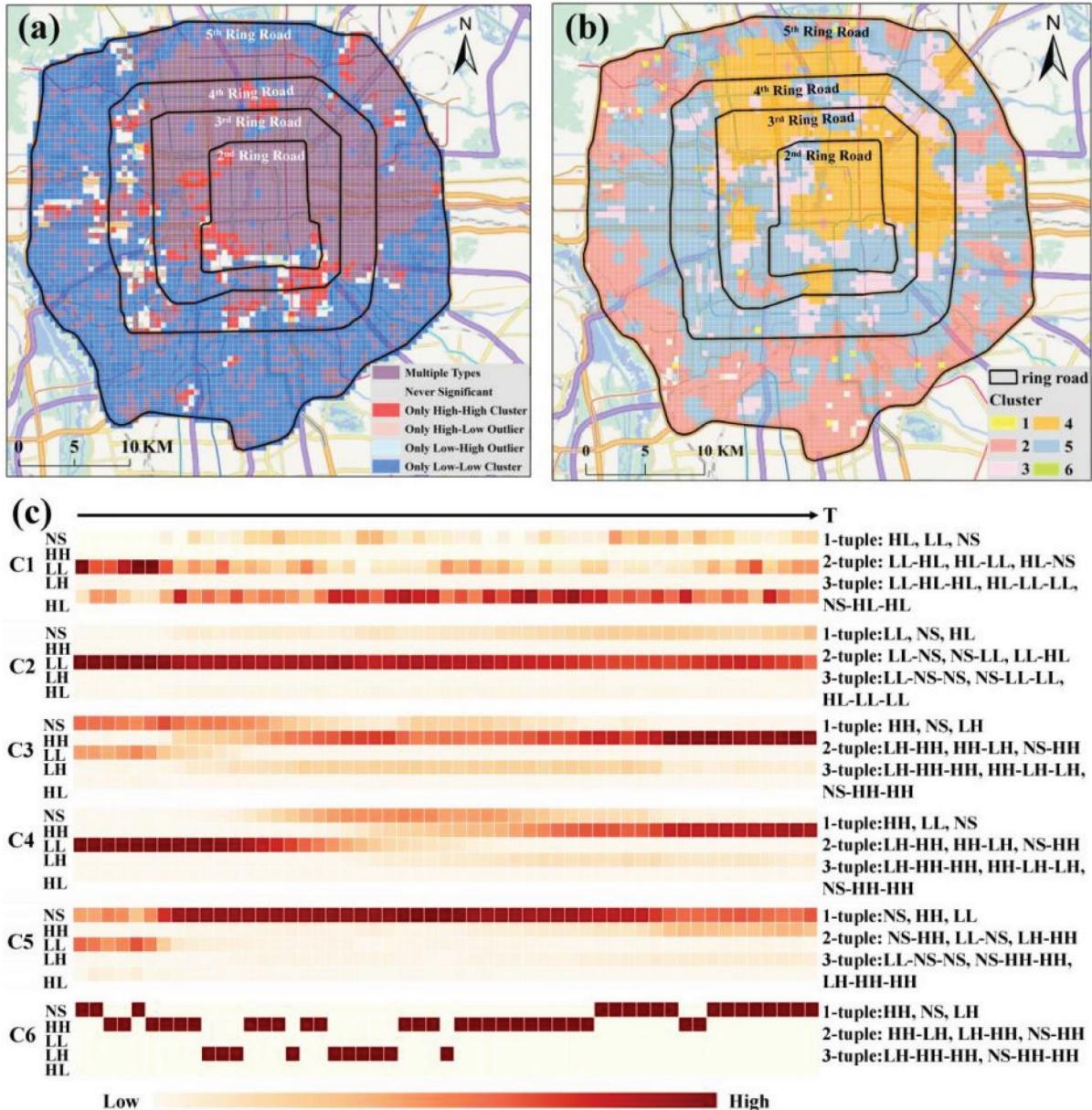


Figure 9. Comparison of evolving S-T association patterns of check-in records using LISA sequence and local outlier analysis in ArcGIS Pro (a) Result of local outlier analysis. (b) Clustering result based on LISA sequence. (c) Heat map and frequent patterns of each category.

simplification of time-varying patterns, failing to capture more intricate, gradual, or subtle changes effectively. In contrast, the proposed LISA sequence approach enables a more detailed characterization of S-T association changes. Figure 9(b-c) illustrates the spatial distribution and frequent patterns of clustering results based on LISA sequence. Figure 9(c) visualizes the sum of the five LISA modes within each cluster over time using heat maps, and the top three frequent subsequences of length 1 to 3 are shown on the right side of each heat map. In the study area, 43.3% of the

sequences fell into category 5, which exhibited a transition from LL in the beginning, to HH in the final period, and insignificant in other time periods. Additionally, Class 3 (474 sequences) and Class 4 (942 sequences) were mainly distributed in the northeast part of the study area, with the most common 3-tuple LH-HH-HH. The proposed method also detected anomalous evolutionary patterns of S-T association in cluster 1 and cluster 6. Cluster 1 contained sequences characterized by alternating LL and HL, with LL-HL as the most frequent 2-tuple. Cluster 6 exhibited a prevailing HH-

LH alternation, with HH-LH as the most common 2-tuple. These two clusters had low similarity to all other sequences, suggesting the presence of anomalous S-T association patterns.

To further interpret the results, the correlation between the evolving S-T association patterns of check-in data and urban land use was analyzed. The S-T association dynamics of human activities reflect how people interact with the different place, which is inherently linked to the land use of the place, such as residential, commercial or industrial. Figure 10 presents the percentage of the clustering categories for each urban land use type. Industrial and transportation zones were dominated by category 2, with category 2 accounting for 68.1% and 56.9% of all categories, respectively. The results indicate that these zones were dominated by LL patterns throughout the time period, indicating low human activity levels. Additionally, 50.7% of the residential areas were classified as category 5, which was dominated by LL from January to February 2017, HH from October to December 2017, and was not significant in other time periods. This pattern suggests a seasonal variation in human activity, potentially influenced by factors such as holiday periods or social events. For commercial areas, the combined proportion of category 3 and 4 was about 45.0%, indicating that the most frequent mode was HH. This reflects sustained high levels of human activity, aligning with the commercial nature of these zones. The public areas presented complex patterns, which contained the abnormal sequence with alternating HH and LH

activity levels in category 6. This irregular pattern can be attributed to contextual factors, such as temporary activities like concerts, which lead to fluctuating crowd dynamics. To confirm the statistical significance of the observed patterns, we conducted a chi-square test (Chi-Square Statistic = 282.23, $p < 0.01$, Degrees of Freedom = 20). The results indicate that the association between clustering categories and land use types is statistically significant, confirming that these patterns did not arise by chance. The evolution patterns of S-T association extracted in this paper are consistent with the characteristics of their respective land use types.

4.2. Evolving S-T association patterns of carbon emission intensity

Global climate change, driven by carbon emissions, poses a serious challenge to economic and social development (Dong et al. 2018). Understanding the dynamic S-T association patterns of carbon emission intensity provides a scientific foundation for formulating effective policies on energy conservation and carbon emission reduction in China. This case study examines the evolution of carbon emission intensity at the urban scale from 2001 to 2020.

4.2.1. Study area and datasets

As shown in Figure 6(a), China is divided into four different economic macro-regions: the economically well-developed coastal parts in Eastern China, the less-developed Central and Northeastern China, and the developing region of Western China. Considering the availability of data, this study focuses on Eastern, Central and Northeastern China as the experimental sample area.

Carbon emission intensity, which is a key metric for evaluating energy utilization quality and carbon emission efficiency, defined as the amount of carbon emissions generated per unit of GDP. Carbon emission data of prefecture-level cities in China from 2001 to 2020 was obtained from Carbon Emission Accounts and Datasets (CEADs). National GDP data from 2001 to 2020 was sourced from China City Statistical Yearbook. Cities with incomplete data were excluded to ensure the accuracy and consistency of our results, resulting in a final dataset of 200 cities. Figure 6(c) shows the carbon emission intensity of 200 cities during 20 years.

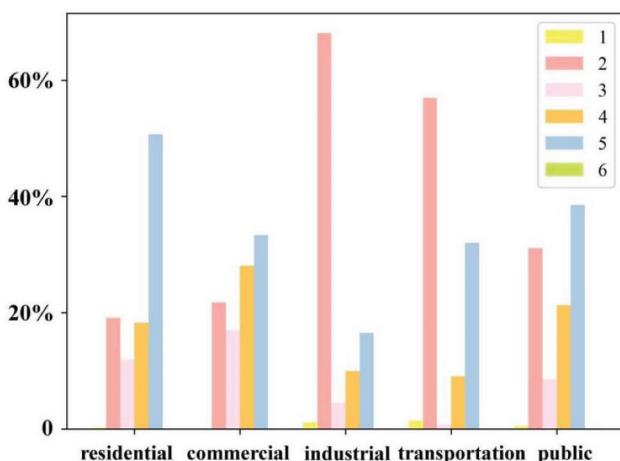


Figure 10. Correlation between S-T association of check-ins and urban land use.

4.2.2. LISA sequence analysis of carbon emission intensity

Figure 11(a) displays the S-T LISA cube of the study area. The LISA sequences at the marked position in Figure 11(a) were selected for visualization, and the proportion of the five modes in each sequence was counted, as shown in Figure 11(b). Among a total of 200 LISA sequences of cities in the research area, there were 58 simple sequences and 142 composite sequences. Simple sequences comprised 29% of the total sequences, with the LL mode being the dominant type (57 sequences). Meanwhile, composite sequences constituted 71% of the total sequences, with 19 sequences having dominant types, while the remaining 123 sequences exhibited various patterns without clear dominant types. These findings demonstrate that the study area displays dynamic and complex S-T autocorrelation patterns, necessitating further exploration and analysis.

The enhanced Hamming distance (Section 3.1) was used to measure sequence similarity, where a greater distance indicates lower similarity. The histogram distribution of the Hamming distances is shown in Figure 12(a). Based on the distance matrix, sequences were clustered hierarchically. Following the elbow diagram (Figure 12(b)) and dendrogram visualization (Figure 12(c)), the optimal number of clustering categories was determined as five.

Figure 13(a) compares the Local Outlier Analysis (LOA) results, which classify most regions as multiple types, failing to capture mode transitions over time. In contrast, Figure 13(b-c) visualizes the spatial clustering results and the frequent LISA subsequences,

offering a detailed temporal evolution of carbon emission intensity. The northeastern cities in Heilongjiang Province are classified as the first category, which is dominated by the HH pattern all the time. The relatively low GDP of these cities has led to high carbon emission intensity. Other cities in northeastern China are classified as Category 6, which show HH initially, followed by insignificant in later years. Most cities in eastern and central China are classified as Category 2, the most commonly occurring 3-tuple of which were NS-LL-LL and HH-NS-NS. The carbon emission intensity of these cities realized a transition from HH to LL at 2010, reflecting the impact of green policies introduced that year. Cluster 4 represents cities that experienced a policy-driven reduction in emissions from HH to LL after the 2015 Paris Agreement, notably in the southern regions of China. Three cities in Guangdong Province and two cities in Hubei Province are classified as Category 3, displaying an anomalous pattern of transitioning from HH to HL to LL. This suggests that these cities lagged behind surrounding areas in terms of emission cuts. The proposed method enables automatic identification of dominant evolving S-T association patterns as well as irregular outliers.

5. Discussion

S-T LISA Sequence serves as the foundation of this paper, which encodes genes related to the dynamics of S-T association and resembles the role of DNA sequences in geo-related domains. Other data structures can easily draw inferences from LISA sequences, extending the classical LISA to a dynamic context. For

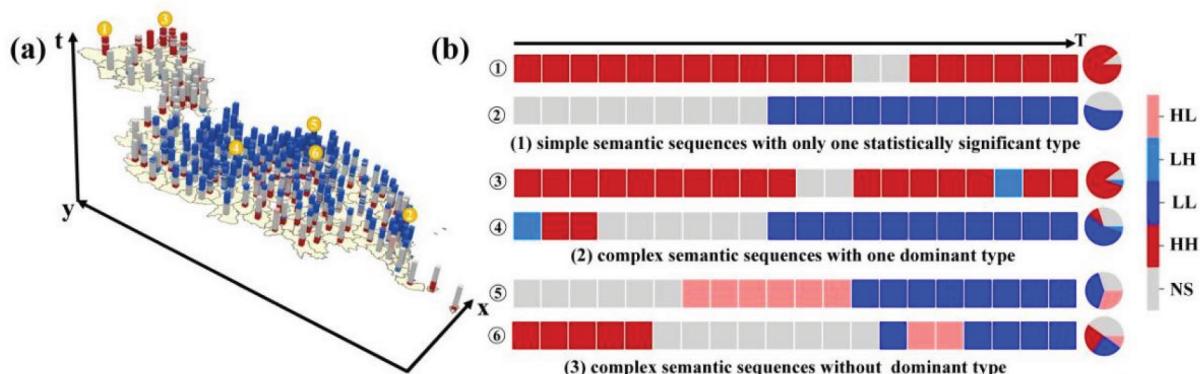


Figure 11. Results of S-T LISA cube and typical LISA sequences of carbon emission intensity. (a) S-T LISA cube of the study area. (b) Typical LISA Sequence.

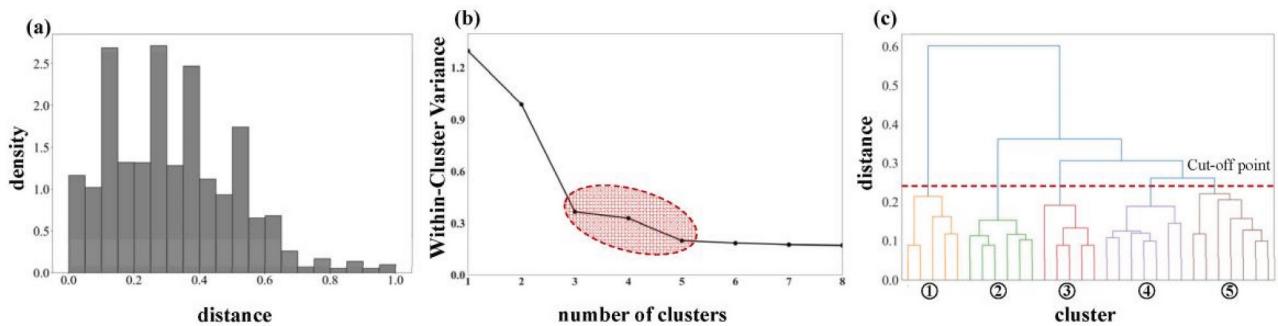


Figure 12. Hierarchical clustering results of carbon emission intensity. (a) Histogram of Hamming distances. (b) Elbow plot for determining the optimal number of clusters. (c) Dendrogram of hierarchical clustering with a cut-off point indicated.

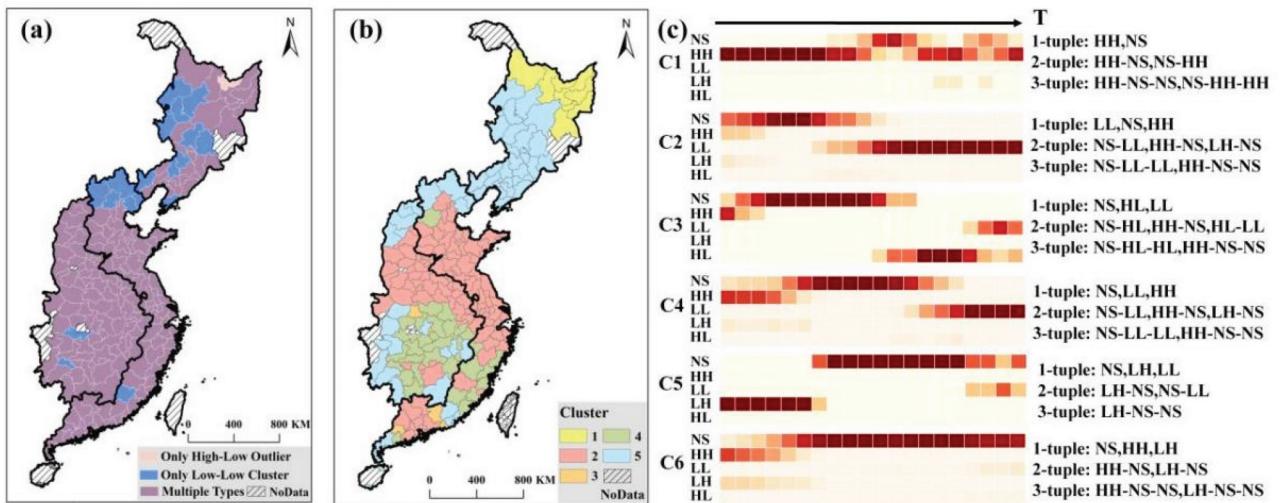


Figure 13. Comparison of evolving S-T association patterns of carbon emission intensity using LISA sequence and local outlier analysis in ArcGIS pro. (a) Result of local outlier analysis. (b) Clustering result based on LISA sequence. (c) Heat map and frequent patterns of each category.

example, by incorporating a time dimension into the Moran scatterplot, a LISA trajectory can be obtained by chronologically connecting all Moran points. Trajectory mining methods (Zheng 2015) can then be applied to analyze S-T LISA trajectories. Furthermore, S-T LISA trajectories can be enriched with type semantics of S-T autocorrelation recorded in LISA sequences, leading to the generation of a LISA semantic trajectory (Parent et al. 2013). These data structures and associated mining methods open up new possibilities for identifying evolutionary patterns in S-T association.

The proposed S-T LISA framework advances traditional S-T cube methods (Putrenko, Pashynska, and Nazarenko 2018) by modeling evolutionary trajectories through LISA sequences and automatically

uncovering hidden patterns via data-driven clustering and frequent subsequence mining. While S-T pattern mining methods based on the S-T cube model rely on fixed classifications, S-T LISA employs a data-driven approach to adaptively reveal hidden evolutionary patterns, such as alternating HH/LL phases or anomalous transitions. These innovations enable S-T LISA to outperform conventional methods in capturing dynamic, complex, and transient S-T association patterns, as demonstrated in applications to carbon emissions and human mobility analysis.

The S-T LISA framework, with its core innovations in dynamic sequence encoding, semantic similarity metrics, and data-driven pattern mining, offers

broad applicability across various domains characterized by spatial-temporal dependencies. In epidemiology and public health (Lin and Wen 2022), the method can track the evolution of disease hotspots and identify sudden outbreaks, informing policies such as vaccination campaigns. Environmental science benefits from its ability to detect recurring ecological changes, such as deforestation (Griffiths et al. 2014) and wildfire spread (Li and Banerjee 2021). Meanwhile, many 3D datasets such as soil data (Lacoste et al. 2014), ocean temperature and salinity dataset (Cheng et al. 2021) at different depths can be applied to our method framework. The LISA sequences are arranged in chronological order for S-T datasets, while they are arranged in depth order for 3D datasets. The framework's flexibility allows it to be adapted to a wide range of fields, making it a powerful tool for spatio-temporal analytics. Future developments could formalize these applications into a domain-agnostic toolkit, further enhancing its potential impact across disciplines.

Several limitations of this work should be noted. First, a fixed time step is set in this method, which overlooks dynamic changes in S-T association at finer temporal granularity. The first case study employed a one-week time step to explore the dynamic population spatial distribution over a year, omitting variations between weekdays and weekends. In the second case study, a one-year time step was employed to explore the evolving S-T association patterns of carbon emission intensity over a span of 20 years, disregarding seasonal variations. To address this, future work will integrate adaptive time-step selection into the S-T LISA framework to offer a multi-scale solution by dynamically adjusting temporal resolution based on the rate of change in spatial associations. Coarser time steps are selected during stable periods, while finer steps are used during periods of rapid changes. This ensures that critical transitions, such as sudden hotspot formation or policy-driven shifts, are accurately detected. For example, in carbon emission analysis, adaptive steps can focus on quarterly intervals during policy implementation years while reverting to annual steps during stable periods. Second, each cluster of S-T association patterns is not merely a statistical grouping but represents meaningful socio-economic or environmental trends. Although external data (e.g. land use maps and policy timelines) have been integrated to enrich the interpretability of these clusters,

further efforts are needed to link the patterns to specific causal mechanisms and validate them through empirical data. In future studies, methods such as regression analysis (Huang, Wu, and Barry 2010) and agent-based modeling (Parker et al. 2003) can be used to test the relationship between cluster membership and factors like economic growth or policy interventions. Another limitation of our proposed method is its computational complexity, particularly when handling large-scale datasets. As the dataset size increases, scalability becomes a critical concern, necessitating optimization to improve processing efficiency and reduce computation time. For instance, when constructing S-T LISA Sequences, each permutation is independent and could be parallelized across CPUs or GPUs to accelerate sequence generation. Additionally, alternative clustering algorithms, such as Self-Organizing Maps (Vesanto and Alhoniemi 2000), could be explored to further optimize performance and enhance scalability.

6. Conclusions

This paper proposes a processing framework for S-T LISA sequences to explore temporal shift patterns of S-T autocorrelation. Considering the finite, orderly, and isometric nature of LISA sequences, an enhanced Hamming distance measure is introduced to quantify their similarity. Based on this measure, LISA sequences with similar characteristics are clustered using the agglomerative hierarchical clustering algorithm, and frequent subsequences in each cluster are identified using the improved FP-Growth algorithm. Two case studies, using grid-scale social media check-in records and city-scale carbon emission data, demonstrate the method's effectiveness in automatically identifying dominant evolutionary patterns and detecting irregular outliers in S-T associations. To the best of our knowledge, this study is the first one to define, characterize, and classify S-T LISA sequences for mining dynamic patterns of S-T association. Integrating sequence pattern mining methods with LISA statistics presents a promising method for uncovering evolving S-T association patterns.

Future research will focus on enhancing the scalability, flexibility, and applicability of the S-T LISA framework. First, multi-scale temporal analysis will be developed through adaptive time-step selection algorithms that dynamically adjust granularity based on data

volatility. Second, alternative clustering algorithms, such as Self-Organizing Maps (SOM), could be explored to further optimize performance and enhance scalability, enabling more efficient and adaptive pattern detection in large datasets. In addition to these methodological advancements, the S-T LISA framework holds significant potential for cross-domain applications. Expanding its use to a broader range of environmental and socio-economic variables could provide deeper insights into dynamic spatial processes and their underlying drivers.

Acknowledgments

We thank the anonymous referees for their helpful and insightful comments and suggestions on this manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Key Research and Development Program of China [grant number 2022YFB3904102], National Natural Science Foundation of China [grant number 42401524] and China National Postdoctoral Program for Innovative Talents [grant number BX20230360].

Data availability statement

The source code and data that support the findings in this study are available at <https://figshare.com/s/19d0c2a8f1d0b1b96532>.

References

- Anselin, L. 1995. "Local Indicators of Spatial Association—LISA." *Geographical Analysis* 27 (2): 93–115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>.
- Berglund, S., and A. Karlström. 1999. "Identifying Local Spatial Association in Flow Data." *Journal of Geographical Systems* 1 (3): 219–236. <https://doi.org/10.1007/s101090050013>.
- Chen, L., and W. Liu. 2011. "An Algorithm for Mining Frequent Patterns in Biological Sequence." In *2011 IEEE 1st International Conference on Computational Advances in Bio and Medical Sciences (ICCBABS)*, edited by I. Mandoiu, 63–68. Orlando, FL, USA: IEEE. <https://doi.org/10.1109/ICCBABS.2011.5729943>.
- Cheng, L., J. Abraham, K. E. Trenberth, J. Fasullo, T. Boyer, R. Locarnini, B. Zhang, et al. 2021. "Upper Ocean Temperatures Hit Record High in 2020." *Advances in Atmospheric Sciences* 38 (4): 523–530. <https://doi.org/10.1007/s00376-021-0447-x>.
- Cheng, S. 2020. "A High-Resolution Emissions Inventory and Its Spatiotemporal Pattern Variations for Heavy-Duty Diesel Trucks in Beijing, China." *Journal of Cleaner Production* 14:119445. <https://doi.org/10.1016/j.jclepro.2019.119445>.
- Dong, F., B. Yu, T. Hadachin, Y. Dai, Y. Wang, S. Zhang, and R. Long. 2018. "Drivers of Carbon Emission Intensity Change in China." *Resources, Conservation & Recycling* 129:187–201. <https://doi.org/10.1016/j.resconrec.2017.10.035>.
- Gao, Y., J. Cheng, H. Meng, and Y. Liu. 2019. "Measuring Spatio-Temporal Autocorrelation in Time Series Data of Collective Human Mobility." *Geo-Spatial Information Science* 22 (3): 166–173. <https://doi.org/10.1080/10095020.2019.1643609>.
- Griffith, D. A., and J. H. P.aelinck. 2018. *Morphisms for Quantitative Spatial Analysis*. Cham: Springer International Publishing.
- Griffiths, P., T. Kuemmerle, M. Baumann, V. C. Radeloff, I. V. Abrudan, J. Lieskovsky, C. Munteanu, K. Ostapowicz, and P. Hostert. 2014. "Forest Disturbances, Forest Recovery, and Changes in Forest Types Across the Carpathian Ecoregion from 1985 to 2010 Based on Landsat Image Composites." *Remote Sensing of Environment* 151:72–88. <https://doi.org/10.1016/j.rse.2013.04.022>.
- Gui, B., A. Bhardwaj, L. Sam, W. Sun, Y. Wang, Y. Yang, J. Yang, et al. 2024. "Revealing the Evolution of Spatiotemporal Patterns of Urban Expansion Using Mathematical Modelling and Emerging Hotspot Analysis." *Journal of Environmental Management* 364:121477. <https://doi.org/10.1016/j.jenvman.2024.121477>.
- Guo, J., H. Zhang, X. Ye, H. Wang, Y. Yang, and G. Tang. 2024. "Spatial Association Measures for Time Series with Fixed Spatial Locations." *International Journal of Geographical Information Science*: 1–25. <https://doi.org/10.1080/13658816.2024.2445185>.
- Hoffmann, L., L. Gilardi, M.-T. Schmitz, T. Erbertseder, M. Bittner, S. Wüst, M. Schmid, and J. Rittweger. 2024. "Investigating the Spatiotemporal Associations Between Meteorological Conditions and Air Pollution in the Federal State Baden-Württemberg (Germany)." *Scientific Reports* 14 (1): 5997. <https://doi.org/10.1038/s41598-024-56513-4>.
- Huang, B., B. Wu, and M. Barry. 2010. "Geographically and Temporally Weighted Regression for Modeling Spatio-Temporal Variation in House Prices." *International Journal of Geographical Information Science* 24 (3): 383–401. <https://doi.org/10.1080/13658810802672469>.
- Jesri, N., A. Saghafipour, A. Koohpaei, B. Farzinnia, M. K. Jooshin, S. Abolkheirian, and M. Sarvi. 2021. "Mapping and Spatial Pattern Analysis of COVID-19 in Central Iran Using the Local Indicators of Spatial Association (LISA)." *BMC Public Health* 21 (1): 1–10. <https://doi.org/10.1186/s12889-021-12267-6>.
- Lacoste, M., B. Minasny, A. McBratney, D. Michot, V. Viaud, and C. Walter. 2014. "High Resolution 3D Mapping of Soil Organic Carbon in a Heterogeneous Agricultural

- Landscape." *Geoderma* 213:296–311. <https://doi.org/10.1016/j.geoderma.2013.07.002>.
- Lee, J., and S. Li. 2017. "Extending Moran's Index for Measuring Spatiotemporal Clustering of Geographic Events." *Geographical Analysis* 49 (1): 36–57. <https://doi.org/10.1111/gean.12106>.
- Li, S., and T. Banerjee. 2021. "Spatial and Temporal Pattern of Wildfires in California from 2000 to 2019." *Scientific Reports* 11 (1): 8779. <https://doi.org/10.1038/s41598-021-88131-9>.
- Lin, C.-H., and T.-H. Wen. 2022. "How Spatial Epidemiology Helps Understand Infectious Human Disease Transmission." *Tropical Medicine and Infectious Disease* 7 (8): 164. <https://doi.org/10.3390/tropicalmed7080164>.
- Liu, L., R. Wang, W. W. Guan, S. Bao, H. Yu, X. Fu, and H. Liu. 2022. "Assessing Reliability of Chinese Geotagged Social Media Data for Spatiotemporal Representation of Human Mobility." *ISPRS International Journal of Geo-Information* 11 (2): 145. <https://doi.org/10.3390/ijgi11020145>.
- Ma, D., T. Osaragi, T. Oki, and B. Jiang. 2020. "Exploring the Heterogeneity of Human Urban Movements Using Geo-Tagged Tweets." *International Journal of Geographical Information Science* 34 (12): 2475–2496. <https://doi.org/10.1080/13658816.2020.1718153>.
- Parent, C., S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M. L. Damiani, et al. 2013. "Semantic Trajectories Modeling and Analysis." *ACM Computing Surveys* 45 (4): 1–32. <https://doi.org/10.1145/2501654.2501656>.
- Parker, D. C., S. M. Manson, M. A. Janssen, M. J. Hoffmann, and P. Deadman. 2003. "Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review." *Annals of the Association of American Geographers* 93 (2): 314–337. <https://doi.org/10.1111/1467-8306.9302004>.
- Putrenko, V., N. Pashynska, and S. Nazarenko. 2018. "Data Mining of Network Events with Space-Time Cube Application." In *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, 79–83. New York: IEEE. <https://doi.org/10.1109/DSMP.2018.8478437>.
- Rey, S. J. 2010. "Spatial Empirics for Economic Growth and Convergence." *Geographical Analysis* 33 (3): 195–214. <https://doi.org/10.1111/j.1538-4632.2001.tb00444.x>.
- Song, W., C. Wang, W. Chen, X. Zhang, H. Li, and J. Li. 2020. "Unlocking the Spatial Heterogeneous Relationship Between per Capita GDP and Nearby Air Quality Using Bivariate Local Indicator of Spatial Association." *Resources, Conservation & Recycling* 160:104880. <https://doi.org/10.1016/j.resconrec.2020.104880>.
- Tao, R., and Y. Chen. 2023. "Applying Local Indicators of Spatial Association to Analyze Longitudinal Data: The Absolute Perspective." *Geographical Analysis* 55 (2): 225–238. <https://doi.org/10.1111/gean.12323>.
- Tao, R., Y. Chen, and J.-C. Thill. 2023. "A Space-Time Flow LISA Approach for Panel Flow Data." *Computers, Environment and Urban Systems* 106:102042. <https://doi.org/10.1016/j.compenvurbsys.2023.102042>.
- Trivedi, R., and H. A. Nagarajaram. 2020. "Substitution Scoring Matrices for Proteins - an Overview." *Protein Science* 29 (11): 2150–2163. <https://doi.org/10.1002/pro.3954>.
- Vesanto, J., and E. Alhoniemi. 2000. "Clustering of the Self-Organizing Map." *IEEE Transactions on Neural Networks* 11 (3): 586–600. <https://doi.org/10.1109/72.846731>.
- Wang, Z., and N. S. N. Lam. 2020. "Extending Getis–Ord Statistics to Account for Local Space–Time Autocorrelation in Spatial Panel Data." *Professional Geographer* 72 (3): 411–420. <https://doi.org/10.1080/00330124.2019.1709215>.
- Wu, B., R. Li, and B. Huang. 2014. "A Geographically and Temporally Weighted Autoregressive Model with Application to Housing Prices." *International Journal of Geographical Information Science* 28 (5): 1186–1204. <https://doi.org/10.1080/13658816.2013.878463>.
- Xu, B., B. Qi, K. Ji, Z. Liu, L. Deng, and L. Jiang. 2022. "Emerging Hot Spot Analysis and the Spatial–Temporal Trends of NDVI in the Jing River Basin of China." *Environmental Earth Sciences* 81 (2): 55. <https://doi.org/10.1007/s12665-022-10175-5>.
- Ye, X., and S. Rey. 2013. "A Framework for Exploratory Space-Time Analysis of Economic Data." *Annals of Regional Science* 50 (1): 315–339. <https://doi.org/10.1007/s00168-011-0470-4>.
- Zhang, H., X. Zhou, Y. Yang, H. Wang, X. Ye, and G. Tang. 2024. "Advancing Process-Oriented Geographical Regionalization Model." *Annals of the American Association of Geographers* 114 (10): 2388–2413. <https://doi.org/10.1080/24694452.2024.2380893>.
- Zheng, Y. 2015. "Trajectory Data Mining." *ACM Transactions on Intelligent Systems and Technology* 6 (3): 1–41. <https://doi.org/10.1145/2743025>.